SEGMENTATION AND ANALYSIS OF TANIĀVARTANAM IN CARNATIC MUSIC CONCERTS

Gowriprasad R[1] **Srikrishnan S**[2] **R Aravind**[1] **Hema A Murthy**[1]
[1] Indian Institute of Technology Madras, [2] Carnatic Percussionist, [1,2]Chennai, India
ee19d702@smail.iitm.ac.in, aravind@ee.iitm.ac.in, hema@cse.iitm.ac.in
[2]srikrishnansridharan@gmail.com

ABSTRACT

In Carnatic music concerts, taniāvartanam is a solo percussion segment that showcases intricate and elaborate extempore rhythmic evolution through a series of homogeneous sections with shared rhythmic characteristics. While taniāvartanam segments have been segmented from concerts earlier, no effort has been made to analyze these percussion segments. This paper attempts to further segment the taniāvartanam portion into musically meaningful segments. A taniāvartanam segment consists of an abhiprāya, where artists show their prowess at extempore enunciation of percussion stroke segments, followed by an optional korapu, where each artist challenges the other, and concluding with mohra and korvai, each with its own nuances. This work helps obtain a comprehensive musical description of the taniāvartanam in Carnatic concerts. However, analysis is complicated owing to a plethora of tāla and naḍe. The segmentation of a taniāvartanam section can be used for further analysis, such as stroke sequence recognition, and help find relations between different learning schools. The study uses 12 hours of taniāvartanam segments consisting of four tāla-s and five naḍe-s for analysis and achieves 0.85 F1-score in the segmentation task.

1. INTRODUCTION

Carnatic music (CM) is a South Indian music tradition considered an ancient form of Indian art music (IAM). A typical CM concert features a lead artist, typically a vocalist, accompanied by a violinist and percussion instrument artists. The lead percussion instrument in this ensemble is usually the *mridangam*, while additional percussion instruments like the *ghatam*, *khanjira*, and *morsing* may also be present. A CM concert includes a solo percussion performance known as *taniāvartanam*, or *tani* for short. Tani is a structured sequence of rhythmic elaborations performed at a fixed metric tempo and bound to a metric cycle (*tāla*). This study attempts to study the elaborations in tani, segment them using a culture-specific approach, and assigns semantically meaningful labels.

Audio recordings of concert performances available online often lack detailed metadata and annotations regarding section boundaries and other information, particularly in the context of IAM. With the increasing availability of music collections and digital devices, there is growing interest in accessing music based on its characteristics. The paucity of editorial metadata has necessitated the development of music information retrieval (MIR) techniques to extract music's characteristic properties from audio recordings automatically. The paper is organized as follows. The taniāvartanam structure is described, followed by the task objectives, challenges, and dataset description. Domain-specific feature engineering is done, and the task is addressed for different cases. The experimental results are analyzed and discussed with culture-specific explanations.

1.1 Taniāvartanam Description

The tani is a highly structured and elaborate percussion performance that is a prominent feature of CM, showcasing the rhythmic skills and creativity of the percussionist. The main percussion instrument is the mridangam, occasionally accompanied by ghatam (clay pot), khanjira, and morsing (Jew's Harp). Since tani is part of a main item, it is performed in the same tāla, and metrical tempo as the main item. The intricacies are based on the precise mathematical calculations of the metric cycle.

The duration of the tani is divided among the mridangam and accompanying percussion to showcase individual artistry, e.g., if mridangam and ghatam are present, the structural framework of the tani is typically as follows: The mridangam always starts first by playing *sarvalaghu* (SV) patterns (indicators of basic tāla structure), and the complex patterns are introduced gradually. These elaborations are performed in a particular rhythm structure called *naḍe* (usually in *chaturaśra* at first) for a few rhythm cycles. These elaborations on a particular rhythmic theme are termed as *abhiprāya*. The literal meaning is "opinion", i.e, the artists' viewpoint of that particular rhythm structure. Ghatam follows and tries to keep the same theme built by the mridangam in the first cycle by playing in the same naḍe [1]. In the second cycle, the mridangist usually may change the naḍe (to *tiśra*, for example) and elaborates. The ghatam usually follows in the same naḍe or switches to a different naḍe (khanḍa). These may or may not continue for more than two cycles, usually owing to time constraints. Each abhiprāya ends with a pattern called *korvai*, which is repeated thrice to arrive at downbeat.

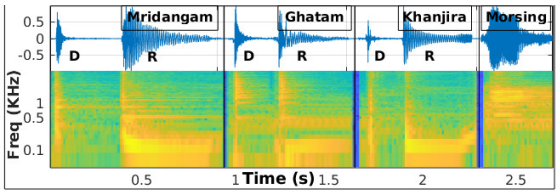

Figure 1. Spectral Illustration of a few Carnatic Percussion Strokes. D: Damped Strokes, R: Resonant Strokes

These abhiprāya-s are followed by the *korapu*, usually seen as a question-answer between mridangam and ghatam. Here it starts with multiple cycles of rhythmic patterns by the mridangam followed by ghatam, where each artist challenges the other. The duration of the rhythm patterns in korapu keeps reducing progressively from full cycle, half cycle, quarter cycle until it finally reduces to a single beat. It can be translated as "rhythmic descent" or "step-by-step reduction". The artist(s) then start playing together playing faster with crisp strokes (*farans*), building up the necessary momentum for playing the last parts of the tani called *mohra* and longer korvai [1]. Each of these has a specific composition structure upon which the artist builds. This structure holds even if only the mridangam is present, except that the korapu part might be absent. Summarizing the sequence of sections in a *tani* segment can be listed as sarvalaghu patterns abhiprāya in a specific naḍe → change of naḍe → back to starting naḍe → korapu → farans → mohra → final korvai [2].

1.1.1 Aspects of timbre and spectral differences among the Carnatic percussion

In Indian music tradition, accompanying instruments are relatively tuned according to the main melodic instrument or voice. The percussion instruments are also categorized on the sonic aspect. Figure 1 illustrates the damped (D) strokes and resonant (R) strokes of Carnatic percussion instruments. Two-sided percussion, mridangam has both the low-frequency and mid-frequency spectra covered. The ghatam occupies a little over the mid-frequency band, and morsing predominantly spreads over the high-mid frequency spectrum and has a larger resonance. Khanjira occupies a low-frequency spectrum a bit less than the left side of the mridangam. This explains the aesthetic quality of the percussion instruments that have been traditionally in use for CM concerts. The tonal nature also enhances the entire concert when played harmoniously.

1.2 Problem Objective and Challenges

This work addresses three primary tasks: (1) Diarization of the audio into mridangam, khanjira, and ghaṭam sections when multiple instruments are present, (2) Estimation of section boundaries using musical attributes, and (3) Classification of segments into broad categories such as abhiprāya, korapu, farans, mohra, and korvai. To achieve these goals, the paper applies techniques from well-researched music genres while also considering the culture-specific characteristics of tani. To improve readability and clarity, several terms are defined in Table 1.

Identifying and understanding the segments in tani is difficult for most CM audiences, except for professionally trained and practicing percussionists. However, this challenge can be addressed if we have a reliable system that can classify the primary segments in tani from audio recordings. Such a system would not only aid in appreciating the art form for a broader audience, but also serve as a valuable learning tool for beginner-level percussion students.

Segment	Audio fragment between any two adjacent detected boundaries that may or may not cover a complete section.
Section	A primary portion of the taniāvartanam. A section can contain multiple compositions and multiple segments.
Naḍe	A modifier to tāla that decides the number of strokes per beat, The subdivision structure within a beat in CM Chaturaśra, Tiśra, Khanḍa are different kinds of naḍe-s
Abhipraya (AB)	A rhythmic elaboration in a particular naḍe during tani.
Korapu (KP)	A musical dialogue between the musicians during performance.
Farans (FA)	The first part of the conclusion in tani where the percussionists play fast to gain momentum toward the end.
Mohra (MO)	Popular rhythmic structure played after the farans hinting the climax of taniāvartanam.
Korvai (KO)	Stroke patterns that are played three times, concluding the tani.

Table 1. Definitions of terms relevant to this paper

Coming to the challenges, tani is very diverse and extempore. The number of percussions may vary across the concerts. The duration of the tani also varies, influencing the number of possible segments. Additionally, the presence of the korapu section is contingent on the number of percussions, which is rare when only mridangam is played. Each rhythmic structure is presented at multiple speeds. This is reflected in the boundary within a single abhiprāya due to sudden tempo changes. The rendition also has small pauses, which may be part of the rhythmic elaboration or due to the artist's presentation style. As a result, the tani segmentation task presents unique challenges to existing audio segmentation methods. Listening to the entire audio carefully to mark the segment boundaries is time-consuming. This underscores the need to develop systems for automatic segmentation and annotations.

1.3 Dataset Description

Experimenting with various shades of tani requires a diverse collection of annotated audio data. As there is no properly annotated dataset available for this task, we collected diverse recordings of tani and labeled them. All the audio data used in this work is a subset of the Charsur Carnatic [3,4], Sangeethapriya [5] datasets along with two audios from [6]. The tani part from the main concerts is extracted by marking the start and end points. Professional performers listened and annotated the boundaries of primary sections in the tani. By doing so, we collected around 12 hours of annotated tani audio. The duration of each tani in the dataset ranges from 6 minutes to 29 minutes, with 11 minutes of mean duration.

The dataset details are described in Table 2. The considered audios comprises of tani played in four major tāla-s of CM [7, 8], namely ādi, miśra chāpu, khanḍa chāpu, and rupaka. The annotations consist of tāla labels, boundary instances, and labels of primary sections of tani. The multiple percussion audios considered in this work have only two instruments along with additional labelings of the instrument name for their respective segments. The dataset is heterogeneous with artist variability (22 mridangam, >12 ghatam, >8 khanjira), tonic, and tempo variability.

	No. of Abhiprāya	No. of Concerts	Duration ~ (hrs:mins)
Mridangam	51	16	02:24
Mrid + Ghat	86	18	04:56
Mrid + Khanj	94	21	05:47
Total	231	55	12:08

Table 2. Dataset Description.

1.4 Related Work

Segmentation and metadata labeling of a music recording have a fairly good research history both in Western [9–11] and IAM traditions [12, 13]. Various acoustic and temporal parameters were used for the segmentation task [9, 14]. Foote et al. [15] proposed a self-distance matrix method to determine the boundary between contrasting musical characteristics. The changes in musical features in Pop and Rock music were used to train the boosted decision stump [16]. Lately, [17] explored neural networks for structural segmentation, spanning various genres [18].

In the context of IAM, different approaches were explored for segmenting the main concert audios in the Dhrupad [13, 19, 20], Hindustani [21, 22], and Carnatic [12, 23–25] music traditions. For instrumental concerts, Vinutha et al. [22] considered the segmentation of sitar and sarod concerts using reliable tempo detection [26]. The analysis of rhythm/percussion in IAM has primarily focused on stroke onset detection [27, 28], stroke recognition [6, 29–33], and sequence modeling [34, 35] percussion pattern identification [36]. Ajay Srinivasamurthy [37] worked on tracking the "downbeat," provided the tāla is known. Tani diarization was also attempted in [4]. Further, mridangam artist identification from tani audio was attempted [38]. Parallel to [38], tabla gharānā recognition from the tabla solo was addressed in [39, 40].

Nevertheless, no attempts have been reported on the structural analysis of Indian solo percussion. This paper attempts to include additional meta-information to the tani portion of a concert, where the audio is segmented based on musical attributes. This can help identify the tāla and enable the association of the cycle of strokes with that of the lyrics of the main composition in CM. The outcomes can help in the concert summarization task and for further MIR studies in the field of percussion, which is crucial as it can give insights into the rhythm of the main item of the concert. Combined with works on meter tracking [7], percussion source separation [41], and stroke recognition [6], this could lead to additional metadata that could be important to an ardent listener or performer.

2. AUDIO FEATURE ENGINEERING

The raw concert audios have to be pre-processed for further analysis. Since each concert is unique in the choice of metric tempo, tonic, and compositional structure, the features used should be based on concert-specific characteristics. At the same time, it should scale inter-concert. We address the tasks by computing relevant features considering the culture-specific musicological perspectives. Initially, the raw audio is pre-processed by computing the Hilbert envelope of the linear prediction (LP) residual on the raw audio [27]. Then the onset detection function (ODF) is computed

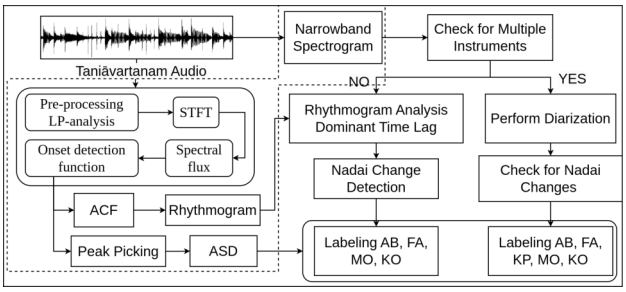

Figure 2. Flow Diagram for Segmentation and Labeling

using the spectral flux method [42]. It is shown to perform on par with state-of-the-art machine learning-based onset detection algorithms on percussion instruments [27]. The computed onset locations are considered for further rhythm analysis. While we have used LP analysis, any onset detection technique could have been used.

2.1 Rhythm and Tempo Features

The change in the rhythm structure or the tempo is a prominent indicator of the section transitions. In the case of percussion instruments, rhythm pattern refers to the aspects of stroke patterns. A rhythm representation can be obtained by considering the stroke ODF (sampled at 10 ms) over a suitably long window and computing the auto-correlation function (ACF). The periodicity analysis using the ACF of the ODF represents the audio in terms of rhythm called rhythmogram [43–45], where rhythm/tempo alone is emphasized.

The ACF of the ODF is computed frame-wise with a frame length of 4 seconds and a frameshift of 0.5 seconds up to a lag of 1 second. The dimension of each frame of the rhythmogram is $p = 100$, corresponding to a 1-second lag. The window length must be large enough to contain sufficient strokes for computing the ACF. Even while playing a slower tempo, we observe at least more than 8-10 strokes (sufficient to calculate the periodicity) in a window length of 4-5 seconds. A uniform window size of 4s is chosen to accommodate variability in rhythm. The peaks along the lag axis of the rhythmogram depict the periodicity of the surface rhythm, indicating surface tempo [22].

The tempo estimation using the product of ACF-DFT [46] is often prone to tempo octave errors due to uneven stroke distribution. We compute the number of strokes in each 4 seconds frame and divide by 4 to get the stroke density at every frame instance. The feature is named average stroke density (ASD), as the averaging is done over 4 seconds frame. The ASD is robust to tempo octave errors and is representative of surface tempo [13]. The mean and std. deviation of strokes per second, as obtained in the entire dataset, are 8.6 and 3.8, respectively. The variance of ASD depicts the tempo diversity in the dataset. Figure 4(c) shows the evolution of ASD over time.

2.2 Spectral Feature

From Section 1.1.1, it is clear that each of the Carnatic percussion instruments has distinct spectral properties, and the spectral features can serve as potential features for instru-

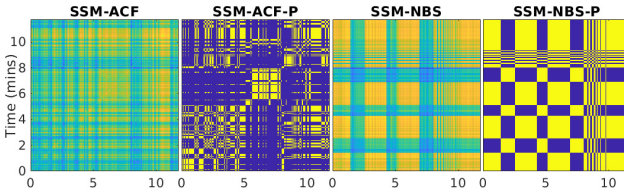

Figure 3. Self-Similarity Matrices on Different Features

ment classification. In this work, we need to localize the segments as coming from one of the percussion. To get the complete spectral aspects of a particular instrument, the spectrum must be computed over a window with almost all kinds of strokes. Thus we computed a narrow band spectrogram (NBS) with a window size of $4s$ and a hop size of $0.5s$. From Section 2.1, we know that the mean ASD is eight strokes per second. Thus in a four-second frame, we can expect at least one resonant stroke. We can clearly distinguish mridangam and ghatam segments from NBS in Figure 4(a).

2.3 Spectral and Rhythm Posteriors

The high-dimensional NBS and ACF rhythmogram represent the spectral and rhythmic-tempo homogeneity within the segment and the changes between the adjacent segments. This allows us to use Gaussian mixture models (GMM) to model the instrument's spectral and temporal homogeneity. The NBS and ACF vectors are converted to spectral and rhythm posteriors (NBS-P, ACF-P), representing class conditional probabilities.

We use two mixtures GMM to represent NBS feature vectors with the intuition that each instrument property is modeled by one mixture. Interestingly we find that each mixture corresponds to a different timbre. We also tried a third mixture to represent the portion where both the instruments play together (FA, MO, KO). This failed due to the volume dominance of mridangam and gave false alarms. The posterior feature computed on NBS is depicted in Figure 4(d). The posteriors from the rhythmogram are computed with five mixture components, each representing a particular speed. The GMM is fit only on the NBS and ACF vectors from a particular concert. The number of Gaussians is determined by the different speeds and naḍe-s expected in a concert.

3. TANI SEGMENTATION AND LABELING

Since tani may contain only mridangam or multiple instruments, we first need to detect if a particular tani audio has multiple instrument or not. The abhiprāya region segmentation task is slightly different in both cases. Locating the abhiprāya boundaries is based on detecting a change in the instrument itself (in case of multiple instrument) and the local rhythmic structure of segments at the highest timescale (in case of solo mridangam). Figure 2 shows the overall steps involved in the task. Each of the segmentation and labeling steps is described here.

3.1 Multiple Instrument Detection

From Sections 1.1.1, we know that different Carnatic percussion instruments differ in their sonic and timbral aspects

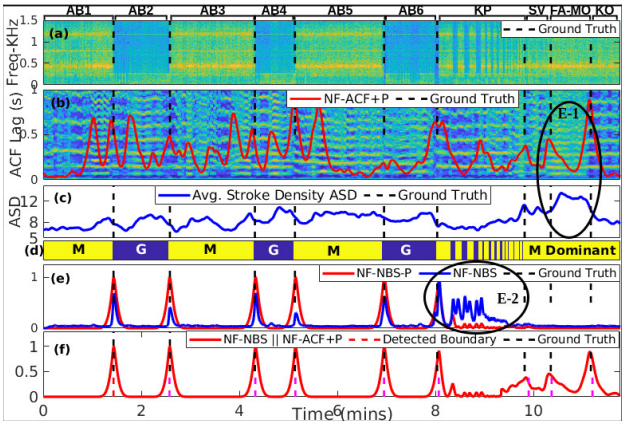

Figure 4. Eg: Multiple Instrument Tani: Segment labels on top (a) NBS feature (b) ACF Rhythmogram with NF-ACF+P overlay-ed (c) ASD evolution over time (d) Posteriors computed on NBS (e) NF-NBS-P (red) obtained from $(15s \times 15s)$ kernel, NF-NBS (blue) from $(3s \times 3s)$ kernel (f) NF-NBS-P replaced with NF-ACF+P in last 2.5 min indicating FA, MO, KO boundaries, and ground truth

and occupy different frequency bins in the spectrum. We use the NBS extracted in Section 2.2 from all the available audios. We built a Gaussian Mixture Model (GMM) on NBS with five mixtures, one for each class – mridangam, ghatam, khanjira, mrid-ghat, mrid-khan. If the ratio of the number of frames from any two classes to the total number of frames in a concert is greater than 20%, then that concert is classified as having multiple instruments. Otherwise, we verify if most frames are from mridangam (at least 80%) and classify it as single instrument mridangam. We performed GMM classification on MFCC features as well. Both methods gave 100% classification accuracy in detecting multiple percussion instruments in a recording.

3.2 Novelty Function Computation

The aim is to get an NF whose peaks indicate the desired segment boundaries. Given the ACF, ACF-P, NBS, and NBS-P feature vectors, the Self-Similarity Matrices (SSM) are computed on each of them using L_2 distance measure [10]. The SSM obtained on the ACF, ACF-P, NBS, and NBS-P are displayed in Figure 3. The homogeneous segments of length L frames possibly appear as $(L \times L)$ blocks. The section change points with high contrast in SSM are captured by convolving a checker-board kernel along the diagonal of SSM [15]. The 1D output obtained is called a novelty function (NF). The peaks of the NF indicate the segment boundary instances having high contrast in SSM. The obtained NFs are (1) the average of NF-ACF, NF-ACF-P (Figure 4(b), Figure 5(a)), (2) NF-NBS, and NF-NBS-P (Figure 4(e).

NFs are computed by convolving $(15s \times 15s)$ kernel with SSM of different features. Peak picking is performed by maintaining a minimum distance between adjacent peaks as 5s. We experimented with smaller kernel sizes such as $(3s \times 3s)$, and $(5s \times 5s)$, resulting in noisy NFs. This decreased the precision due to a lot of false positives. Though much larger kernel sizes, such as $(50s \times 50s)$, made the NFs smoother, they compromised in

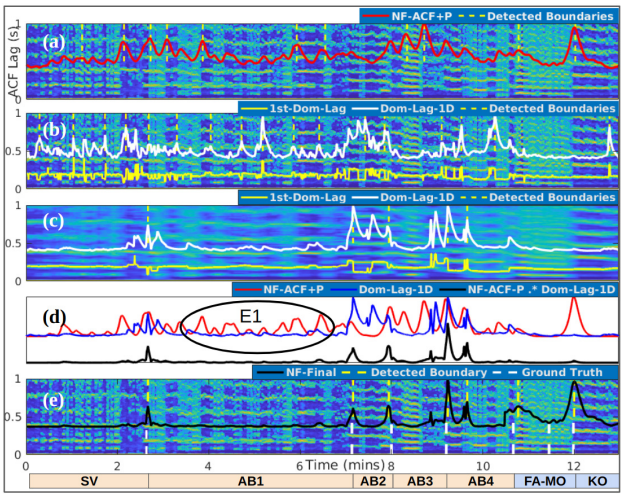

Figure 5. Eg: Solo Mridangam Tani, (a) ACF Rhythmogram with NF-ACF+P overlay-ed along with detected peaks (b) First Dominant peak along the Lag axis FDL (yellow), and its $1st$ diff. FDL-1D highlighting the discontinuities (c) FDL computed on the Gaussian smoothed ACF (yellow) and its $1st$ diff. FDL-1D (white) (d) NF-FDL-ACF (black) is a point-wise product of NF-ACF+P (red) and FDL-1D (blue) (e) NF-FDL-ACF replaced with NF-ACF+P in last 2.5 min indicating FA, MO, KO boundaries along with ground truth, and the segment labels below

resolving the closer boundaries. All the features and NFs in this work are computed at the resolution of 0.5 seconds.

3.3 Case1: Multiple Instrument Tani

In the case of multiple instrument tani, each round of individual percussion elaboration is considered one abhiprāya (one thematic development). Thus instrument change point detection is necessary and sufficient for getting the abhiprāya boundaries. Since the instrument change points are visually evident from the NBS, we used NF-NBS and NF-NBS-P to get the boundaries. A NF obtained from a smaller kernel enhances the rapid instrument change in the KP section, useful in localizing the KP section but creating false positives during segmentation. The first portion of the KP section is fairly large. A larger kernel emphasizes only the start instance of KP by suppressing the rapid instrument change. Thus we used NF obtained from a larger $(15s \times 15s)$ kernel for the segmentation task and the smaller $(3s \times 3s)$ kernel NF for localizing the KP section.

The FA, MO, and KO are always played toward the end of the tani and the FA has a higher ASD. As we see in the Figure 4(e), NF-NBS and NF-NBS-P do not capture these change points. Thus we replace the last two and a half minutes of NF-NBS-P with the average of NF-ACF and NF-ACF-P (NF-ACF+P). This gives the final NF ('red' curve in Figure 4(g)) in the case of multiple instrument tani. We empirically choose the last 2.5 mins as the FA, MO, KO are always found in the last 2.5 mins in the entire dataset.

3.4 Case2: Solo Mridangam Tani

Computing the AB boundaries on solo mridangam tani is a tough task, as the AB change needs to be detected based on the rhythm (naḍe) change. Naḍe change detection is pivotal in getting the AB boundaries, especially in the case of solo mridangam tani. Relying only on the raw rhythmogram features (NF-ACF+P) creates false alarms due to multiple tempo changes and irregularities within a single AB segment. This necessitates the computation of a robust function to tempo octave changes but also captures the non-octave tempo changes that indicate the naḍe changes.

We initially set to track the first peak along the lag axis of the rhythmogram over time, and the change in the peak lag apart from doubling and halving is expected to indicate the naḍe change. But this is also found to be noisy ('yellow' curve in Figure 5(b)). Thus, we perform horizontal Gaussian smoothing on the rhythmogram to mask the irregularities, then pick the first dominant lag peak (FDL). This fetched a smoother curve ('yellow' curve in Figure 5(c)) having discontinuities around the naḍe change with less tempo octave errors. The peaks on the first difference of this curve (FDL-1D) gave fairly good naḍe change estimates, along with a few false positives. We can observe that the peaks of both NF-ACF+P and FDL-1D (Figure 5(d)-E1) coincide around the naḍe change instances but not elsewhere. Thus we perform *"AND"* operation by multiplying NF-ACF+P and FDL-1D to get a NF which is an indicator of naḍe change. We can observe that the false positives are considerably reduced. Again we can see that towards the last FA-MO-KO portion, this NF is not indicating FA-MO-KO boundaries. Thus, we replace the last two and a half minutes of NF-FDL-ACF with NF-ACF+P, similar to Case1. This gives the final NF in the case of solo mridangam tani ('black' curve in Figure 5(e)).

3.5 Section Classification and Labeling

Given the hypothesized segment boundaries, the task is to classify each segment with appropriate labels. Each section, AB, KP, FA, MO, and KO, has unique structural, positional, and duration characteristics common across the concerts. We use the characteristic musical cues to classify and label the segments. For the multiple instrument tani, a NF obtained from a smaller kernel $(3s \times 3s)$ gives multiple peaks in the KP portion. The hypothesized segment having multiple peaks is labeled as KP [Figure 4(e)(E-2)]. The segments before the KP are classified broadly as AB. We compute the mean of ASD in each segment. As the ASD is high during FA-MO, the segment after KP having the highest mean-ASD is labeled FA [Figure 4(c)(E-1)], followed by KO at last. Labeling of FA, MO, and KO is the same for solo mridangam concerts as well. Korapu is not present if only mridangam is present. All the segments before FA are broadly labeled as AB for solo mridangam concerts. Thus the algorithm with a set of rules based on the structure of tani and the domain knowledge performs classification and labeling. Implementation, annotations, and dataset details are shared for research purposes [1] .

4. ANALYSIS OF RESULTS AND DISCUSSION

The tani structural segmentation task is approached as a boundary detection task, where the presence or absence of

[1] https://bit.ly/3XIJfMa

Case	Section	Precision	Recall	F1-Score
Multiple Percussion	AB	0.92	0.99	0.96
	KP-FA-MO-KO	0.82	0.89	0.86
	Overall	0.87	0.94	**0.91±0.03**
Single Percussion	AB	0.7	0.82	0.74
	FA-MO-KO	0.82	0.86	0.83
	Overall	0.75	0.84	**0.79±0.05**

Table 3. Segmentation Results

a boundary is examined in uniformly spaced feature frames of 0.5 seconds. Unlike stroke onset detection, the task is addressed at a larger time scale and thus has a tolerance duration in "seconds" rather than milliseconds [27, 47]. A true-positive detection is one where the prediction boundary falls within ± 3 seconds of the ground truth boundary, while a false-positive detection is one where it does not. Precision, recall, and F1-scores are used for evaluation. Evaluation is performed on the entire dataset, as the proposed method is unsupervised, and no model training is done.

The segmentation evaluation scores for each case and individual sections are tabulated in Table 3. The recall is good in all cases, indicating that the system successfully detects the desired boundaries considerably. We can observe that the precision is consistently less than recall, indicating false positives. The change in local rhythm structure, which may be both gradual and abrupt, causes peaks in the NFs. The gradual change in rhythm structure can be seen often in the AB section as it is extempore.

In Case 1, the AB boundaries are identical to the instrument switching instances, and the NF-NBS/NF-NBS-P captured it well with a 0.96 F1-score. The KP-FA-MO-KO section performance is slightly lower, as the rapid instrument switching caused false positives. The end of the KP section is not always evident as the cycle duration reduces to one beat. A small SV pattern may also exist after KP while moving towards FA, making boundary detection challenging. Since MO is played along with or immediately follows the FA, the FA-MO boundary is often missed, reducing recall.

In Case 2, the AB boundaries are not straightforward. The local variations, tempo doubling and halving cause false positives when the NF-ACF and NF-ACF-P are used. These local variations also cause the first dominant lag on the ACF to be noisy. The horizontal averaging of the rhythmogram aided in noise-free first dominant lag tracing and considerably reduced false positives, but still, the false alarms persisted. The naḍe changes are also very gradual in many cases, which are not evident with tempo-related ACF analysis. For example, while transiting from 6 to 5 strokes per beat, the change is hardly noticeable when the metric tempo is fast. A few of the AB boundaries are also missed during smoothing. The performance on the FA-MO-KO is similar to Case:1, as the NF-ACF+P is used in the last 2.5 mins for both cases. Case 2 has more variance in F1-Score than Case 1. The average F1-score for both cases combined is 0.83.

We also experimented with ± 5s and ± 1s tolerance windows. The overall recall increased by 0.2 with a marginal increment in precision for the ± 5s case. The ± 1s case reported a drop of precision and recall by 0.4 and 0.3, respectively. This is evident as 1s corresponds to only two feature frames in this work, and many boundaries are missed.

Section classification performance is evaluated by considering the ground truth markings. We quantify the performance of calculating the ratio of correctly classified frames to the total number of frames in a tani. The weighted average of correctly classified frames in the entire dataset considering the lengths of each tani is 92%. That is, given 10m of segmented tani, around 9m-15s of the frames are correctly labeled as AB, KP, FA, MO, KO.

5. CONCLUSIONS

This work has addressed an unexplored problem, structural segmentation, and labeling of tani audios. We motivate the problem and present different facets and challenges in the task. From the experiments performed, it is clear that individual features alone are inadequate for segmentation. A culture-specific approach is clearly required, both in feature choice and modeling. Timbre is used when it is required to detect if multiple instruments are present in the tani, and MFCC features were found to be adequate. On the other hand, detecting AB sections required analysis of both timbre and rhythmogram to detect boundaries. Identifying AB sections when two percussion instruments are present is quite easy. In contrast, determining AB sections in a solo percussion instrument is difficult as naḍe changes/speed changes are difficult to determine. The hope is that such a task will aid in including additional meta-data w.r.t a concert.

The major contributions of this work are as follows: (i) curating a diverse dataset of tani recordings of around 12 hours having section boundary information along with primary section labels, (ii) evaluating the existing MIR techniques with culture-specific adaptation for a musicologically important task, segmentation and labeling of tani, (iii) formulating average stroke density (ASD) feature (a representative of surface tempo), which is robust to tempo octave errors, (iv) formulating the class-conditional probability features from the rhythmogram, and spectral features, and (v) exploring the combination of different NFs obtained from different features to achieve the task. Finally, this work provides an example of adapting available MIR methods to genre-specific problems by performing appropriate feature engineering.

6. ACKNOWLEDGMENTS

The authors are grateful to the percussion maestros V Selvaganesh, Patri Satish Kumar and Giridhar Udupa for their support and help. We are thankful to Ajay Srinivasamurthy for his support and timely guidance. We thank Jom Kuriakose for sharing the dataset audios.

7. REFERENCES

[1] U. Giridhar. (2020) Description of tani avartanam. [Online]. Available: https://www.ghatamudupa.com/

[2] E. N. Sunil, *Resounding Mridangam: The Majestic South Indian Drum*. Erickavu N Sunil, March 2021. [Online]. Available: https://www.youtube.com/c/erickavunsunil

[3] Charsur digital workstation. [Online]. Available: https://musicbrainz.org/label/3e188240-9eb5-4842-b7b9-d6c2393211b7

[4] N. Dawalatabad, J. Kuriakose, C. C. Sekhar, and H. A. Murthy, "Information bottleneck based percussion instrument diarization system for taniavartanam segments of carnatic music concerts." in *INTERSPEECH*, 2018.

[5] Sangeethapriya – indian fine arts. [Online]. Available: https://www.sangeethapriya.org/

[6] J. Kuriakose, J. C. Kumar, P. Sarala, H. A. Murthy, and U. K. Sivaraman, "Akshara transcription of mrudangam strokes in carnatic music," in *Twenty First National Conference on Communications (NCC) 2015*.

[7] A. Srinivasamurthy, A. Holzapfel, A. T. Cemgil, and X. Serra, "Particle filters for efficient meter tracking with dynamic bayesian networks," in *Proc. 16th International Society for Music Information Retrieval (ISMIR), Málaga, Spain. Canada*, 2015.

[8] A. Srinivasamurthy, G. K. Koduri, S. Gulati, V. Ishwar, and X. Serra, "Corpora for music information research in indian art music," in *Proc. International Computer Music Conference, ICMC/SMC; Athens, Greece.*, 2014.

[9] R. B. Dannenberg and M. Goto, "Music structure analysis from acoustic signals," in *Handbook of signal processing in acoustics*. Springer, 2008, pp. 305–331.

[10] J. Paulus, M. Müller, and A. Klapuri, "State of the art report: Audio-based music structure analysis." in *Proc. 11th International Society for Music Information Retrieval (ISMIR)*. Utrecht, 2010, p. 625–636.

[11] O. Nieto, "Discovering structure in music: Automatic approaches and perceptual evaluations," Ph.D. dissertation, New York University, 2015.

[12] S. Padi and H. A. Murthy, "Segmentation of continuous audio recordings of carnatic music concerts into items for archival," *Sādhanā*, vol. 43, no. 10, pp. 1–20, 2018.

[13] P. Rao, T. P. Vinutha, and M. A. Rohit, "Structural segmentation of alap in dhrupad vocal concerts," *Transactions of the International Society for Music Information Retrieval*, vol. 3, no. 1, 2020.

[14] P. Grosche, M. Müller, and F. Kurth, "Cyclic tempogram—a mid-level tempo representation for musicsignals," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2010.

[15] J. Foote, "Automatic audio segmentation using a measure of audio novelty," in *Proc. International Conference on Multimedia and Expo. (ICME)*. IEEE, 2000.

[16] D. Turnbull, G. R. Lanckriet, E. Pampalk, and M. Goto, "A supervised approach for detecting boundaries in music using difference features and boosting." in *Proc. 8th International Society for Music Information Retrieval (ISMIR)*, 2007.

[17] K. Ullrich, J. Schlüter, and T. Grill, "Boundary detection in music structure analysis using convolutional neural networks." in *ISMIR*, 2014, pp. 417–422.

[18] J. B. L. Smith, J. A. Burgoyne, I. Fujinaga, D. De Roure, and J. S. Downie, "Design and creation of a large-scale database of structural annotations." in *Proc. 22nd International Society for Music Information Retrieval (ISMIR)*, 2011.

[19] M. A. Rohit and P. Rao, "Structure and automatic segmentation of dhrupad vocal bandish audio," *Unpublished technical report*, 2020.

[20] M. A. Rohit, T. P. Vinutha, and P. Rao, "Structural segmentation of dhrupad vocal bandish audio based on tempo," in *Proc. International Society for Music Information Retrieval (ISMIR)*, 2020.

[21] P. Verma, T. P. Vinutha, P. Pandit, and P. Rao, "Structural segmentation of hindustani concert audio with posterior features," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE*, 2015.

[22] T. P. Vinutha, S. Sankagiri, K. K. Ganguli, and P. Rao, "Structural segmentation and visualization of sitar and sarod concert audio." in *Proc. 17th International Society for Music Information Retrieval (ISMIR)*, 2016.

[23] K. S. PV, S. Sankaran, and H. Murthy, "Segmentation of carnatic music items using kl2, gmm and cfb energy feature," in *Proc. Twenty Second National Conference on Communication (NCC)*. IEEE, 2016.

[24] H. Ranjani and T. Sreenivas, "Hierarchical classification of carnatic music forms," in *Proc. 14th International Society for Music Information Retrieval (ISMIR)*, 2013.

[25] B. Thoshkahna, M. Müller, V. Kulkarni, and N. Jiang, "Novel audio features for capturing tempo salience in music recordings," in *Proc. International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2015.

[26] T. P. Vinutha, S. Sankagiri, and P. Rao, "Reliable tempo detection for structural segmentation in sarod concerts," in *Proc. Twenty Second National Conference on Communication (NCC)*. IEEE, 2016.

[27] R. Gowriprasad and K. S. R. Murty, "Onset detection of tabla strokes using lp analysis," in *Proc. International Conference on Signal Processing and Communications (SPCOM)*. IEEE, 2020.

[28] P. A. M. Kumar, J. Sebastian, and H. A. Murthy, "Musical onset detection on carnatic percussion instruments," in *Proc. Twenty First National Conference on Communications (NCC)*, 2015.

[29] O. Gillet and Richard, "Automatic labelling of tabla signals," in *Proc. 4th International Society for Music Information Retrieval (ISMIR)*, 2003.

[30] P. Chordia, "Segmentation and recognition of tabla strokes." in *Proc. 6th International Society for Music Information Retrieval (ISMIR)*, 2005.

[31] K. Samudravijaya, S. Shah, and P. Pandya, "Computer recognition of tabla bols," Technical report, Tata Institute of Fundamental Research, Tech. Rep., 2004.

[32] M. A. Rohit, A. Bhattacharjee, and P. Rao, "Four-way classification of tabla strokes with models adapted from automatic drum transcription," in *Proc. 22nd International Society for Music Information Retrieval (ISMIR)*, 2021.

[33] A. Anantapadmanabhan, A. Bellur, and H. A. Murthy, "Modal analysis and transcription of strokes of the mridangam using non-negative matrix factorization," in *IEEE International Conference on Acoustics, Speech and Signal Processing, 2013*.

[34] P. Chordia, A. Sastry, and S. Şentürk, "Predictive tabla modelling using variable-length markov and hidden markov models," *Journal of New Music Research*, vol. 40, no. 2, pp. 105–118, 2011.

[35] P. Chordia, A. Sastry, T. Mallikarjuna, and A. Albin, "Multiple viewpoints modeling of tabla sequences." in *Proc. 11th International Society for Music Information Retrieval (ISMIR), 2010*.

[36] S. Gupta, A. Srinivasamurthy, M. Kumar, H. A. Murthy, and X. Serra, "Discovery of syllabic percussion patterns in tabla solo recordings," in *Proc. 16th International Society for Music Information Retrieval (ISMIR); 2015*.

[37] A. Srinivasamurthy, "A data-driven bayesian approach to automatic rhythm analysis of indian art music," Ph.D. dissertation, Universitat Pompeu Fabra, 2017.

[38] K. Gogineni, J. Kuriakose, and H. A. Murthy, "Mridangam artist identification from taniavartanam audio," in *Proc. Twenty Fourth National Conference on Communications (NCC)*. IEEE, 2018.

[39] R. Gowriprasad, V. Venkatesh, H. A. Murthy, R. Aravind, and K. S. R. Murty, "Tabla Gharana Recognition from Audio Music recordings of Tabla Solo performances," in *Proc. 22nd International Society for Music Information Retrieval Conference*, 2021.

[40] R. Gowriprasad, V. Venkatesh, and S. R. Murty K, "Tabla gharana recognition from tabla solo recordings," in *Proc. National Conference on Communications (NCC)*, 2022.

[41] N. Dawalatabad, J. Sebastian, J. Kuriakose, C. C. Sekhar, S. Narayanan, and H. A. Murthy, "Front-end diarization for percussion separation in taniavartanam of carnatic music concerts," *arXiv preprint arXiv:2103.03215*, 2021.

[42] S. Dixon, "Simple spectrum-based onset detection," *MIREX 2006*, p. 62, 2006.

[43] K. Jensen, "Multiple scale music segmentation using rhythm, timbre, and harmony," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–11, 2006.

[44] P. Grosche and M. Muller, "Extracting predominant local pulse information from music recordings," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 6, pp. 1688–1701, 2010.

[45] K. K. Jensen, "Rhythm-based segmentation of popular chinese music," in *Proc. 6th International Society for Music Information Retrieval (ISMIR)*, 2005.

[46] G. Peeters, "Template-based estimation of time-varying tempo," *EURASIP Journal on Advances in Signal Processing*, vol. 2007, pp. 1–14, 2006.

[47] J. P. Bello, L. Daudet, S. Abdallah, C. Duxbury, M. Davies, and M. B. Sandler, "A tutorial on onset detection in music signals," *IEEE Transactions on speech and audio processing*, vol. 13, no. 5, pp. 1035–1047, 2005.